1Г

УДК 004.65:004.89

А.Я. Гладун[1]*, Ю.В. Рогушина*[2]*, Л.В. Петрухина*[1]
[1]Международный научно-учебный центр информационных технологий и систем НАНУ и МОНУ, г. Киев, Украина
[2]Институт программных систем НАНУ, г. Киев, Украина
glanat@yahoo.com, _jjj_@ukr.net

Использование технологии Semantic Web для управления знаниями в системах Business Intelligence

В статье рассматриваются тенденции развития Business Intelligence, его связь с другими направлениями IT, анализируется целесообразность применения в Business Intelligence 2.0 технологий Semantic Web, в частности онтологического анализа и стандартов представления метаданных.

Введение

Сегодня создание и внедрение технологий Business Intelligence (BI) сформировалось в самостоятельное динамично развивающееся направление индустрии информационных технологий. Сам термин «Business Intelligence» используется сравнительно давно, но со временем несколько менялся смысл, вкладываемый в это понятие. В [1] термин «Business Intelligence» обозначает ориентированный на пользователя процесс, который обеспечивает доступ и исследование **информации**, ее анализ, выработку интуитивного понимания, которые ведут к улучшенному и неформальному принятию решений.

Цель BI – превращение данных в знания, а знаний в бизнес-действия для получения выгоды, преимуществ, усиления конкурентоспособности и устойчивости в условиях кризисов. К сожалению, адекватного перевода Business Intelligence на сегодняшний день в русскоязычной литературе нет, а те термины, что используются, либо некорректны, либо являются копией с английского («бизнес-интеллект», «бизнес-анализ», «разведка бизнеса» и пр.) [2].

В русском языке слово «интеллект» однозначно понимается как способность человека к мышлению. Однако корректнее «интеллект» переводится на английский как «intellect», в то же время как английское слово «intelligence» более многозначно и означает в различных контекстах: 1) способность узнавать и понимать; 2) готовность к пониманию; 3) знания, переданные или приобретенные путем обучения, исследования или опыта; 4) действие или состояние в процессе познания; 5) разведку, разведывательные данные.

Возможно, следует обратиться к переводу слова Intelligence в словосочетании Artificial Intelligence, где под Intelligence понимается вовсе не «интеллект» и тем более не «интеллигентность», а просто способность к логическому выводу. Таким образом, предполагается, что словосочетание «Business Intelligence» можно перевести как «средства логического вывода, ориентированные на бизнес-приложения».

В различных источниках встречаем следующие определения BI:

1. Широкий класс приложений и технологий для сбора, хранения, анализа и обеспечения доступа к данным, направленных на то, чтобы помочь работникам предприятия принимать лучшие бизнес-решения.

2. Использование высокоуровневого программного обеспечения для бизнес-приложений, в частности, набор передовых технологий, которые позволяют сделать систему более интеллектуальной.

3. Программное обеспечение, которое позволяет бизнес-пользователям видеть и использовать большие объемы данных сложной структуры.

Проблема этих определений заключается в том, что они учитывают только программные и технологические аспекты, не учитывая человеческий фактор, акцент на который делается в таких определениях Business Intelligence:

1. Дисциплина, посвященная абстрактному пониманию бизнеса.

2. Среда в которой бизнес-пользователи получают и могут анализировать данные, являющиеся понятными, удобными для обработки, непротиворечивыми, достоверными способами.

Обобщая вышеупомянутые определения, можно сказать, что BI – способность предприятия эффективно использовать свои человеческие и информационные ресурсы [3], [4].

Постановка задачи

Сегодня создание и внедрение технологий Business Intelligence сформировалось в самостоятельное, динамически развивающееся и перспективное направление индустрии информационных технологий (IT). Системы Business Intelligence являются наиболее востребованными приложениями IT для бизнеса, технологическим «полигоном», на котором испытываются и отрабатываются новые знание-ориентированные концепции, использующие алгоритмы искусственного интеллекта. Технология Business Intelligence является интегратором, потребителем и сферой применения разнообразных технологий, в том числе Semantic Web.

К сожалению, современные промышленные приложения не всегда своевременно внедряют научные достижения. Поэтому основной задачей данной работы является анализ основных тенденций развития BI-систем, а также стандартов, концепций и средств технологии Semantic Web, которые призваны увеличить эффективность традиционных BI-систем.

Классификация продуктов Business Intelligence

Современное программное обеспечение Business Intelligence можно подразделить на две категории – BI-инструменты и BI-приложения (рис. 1).

Большинство BI-инструментов применяются конечными пользователями для доступа, анализа и генерации отчетов по данным, которые чаще всего располагаются в хранилище, витринах данных или оперативных складах данных [5]. Среди них выделяют генераторы запросов и отчетов; инструменты оперативной аналитической обработки (online analytical processing, OLAP); корпоративные BI-наборы (enterprise BI suites, EBIS); BI-платформы. Средства генерации запросов и отчетов в большой степени поглощаются и замещаются корпоративными BI-наборами. Многомерные OLAP-механизмы или серверы, а также реляционные OLAP-механизмы являются BI-инструментами и инфраструктурой для BI-платформ.

Многомерные OLAP-механизмы или серверы, а также реляционные OLAP-механизмы являются BI-инструментами и инфраструктурой для BI-платформ. OLAP (англ. online analytical processing, аналитическая обработка в реальном времени) – технология обработки информации, включающая составление и динамическую публикацию отчётов и документов.

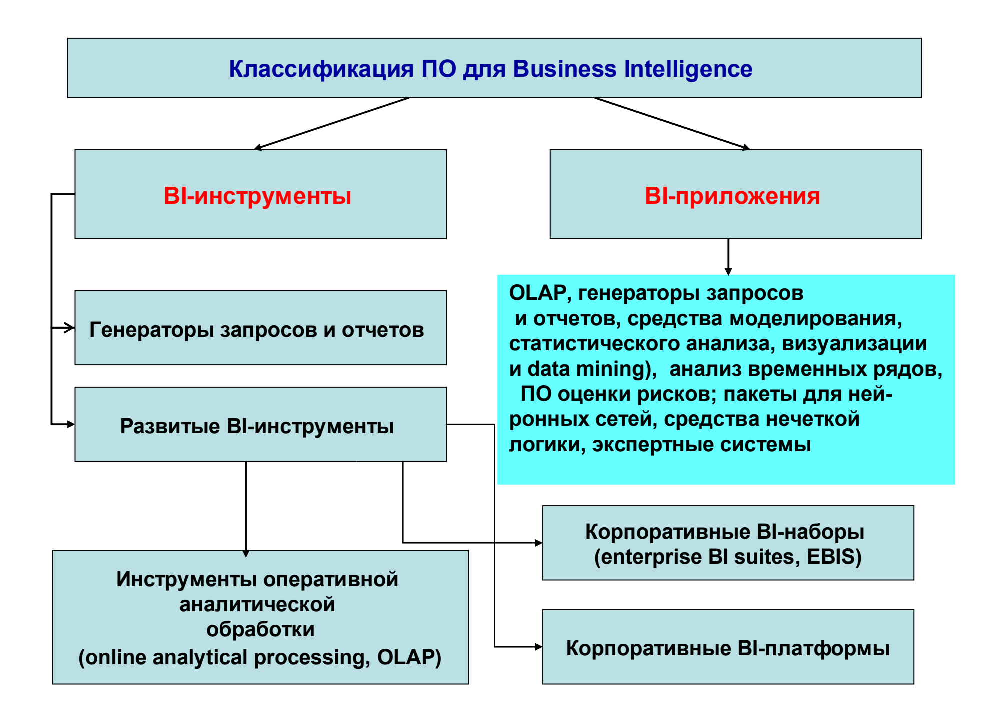


Рисунок 1 – Классификация ПО, используемого в современных BI-системах

Инструменты OLAP являются аналитическими инструментами, которые первоначально были основаны на многомерных базах данных (МБД). МБД сконструированы специально для поддержки анализа количественных данных с множеством измерений (времени, географии, организационных единиц, клиентов, продуктов и т.д.). OLAP позволяет организовать эти измерения в виде иерархии. Данные представлены в виде гиперкубов (кубов) – логических и физических моделей показателей, коллективно использующих измерения, а также иерархии в этих измерениях. Для поддержки МБД используются OLAP-серверы (например, Hyperion Essbase Server) либо эмуляция МБД при помощи реляционных СУБД.

Разработчики приложений используют BI-платформы для создания и внедрения BI-приложений, которые не рассматриваются как BI-инструменты.

Согласно исследованиям ведущих аналитиков [6], [7] ведущие мировые компании на рынке приложений для BI могут быть ранжированы как лидеры – поставщики с широкими функционалом возможностями платформ BI [8], [9] (Cognos, Business Objects, Microsoft, Oracle, MicroStrategy, SAS), претенденты на лидерство (SAP, Information Builders), провидцы (QlikTech, Tibko Spotfire) имеют хорошее представление о платформе BI, отличаются открытостью и гибкостью архитектуры, нишевые игроки (Actuate, Arcplan, Board International, Panorama Software) – компании, достигшие успеха только в определенном сегменте рынка, например, генерации отчетов.

Business Intelligence и системы управления знаниями

Иногда BI определяют как знания, добытые о бизнесе с использованием различных аппаратно-программных технологий, а также способы и методы представления и использования этих знаний. BI позволяет организациям превращать данные в информацию, а затем информацию в знания. При этом достаточно четко разграничивает такие понятия, как «данные», «информация» и «знания». **Данные** понимаются как сведения, которые компьютер записывает, хранит и обрабатывает. Они

не имеют семантической интерпретации. **Информация** – это некие сведения, которые человек имеет об определенной предметной области (ПрО). **Знания** – это информация, которая может использоваться для получения новой информации (правила, закономерности и т.п.).

С каждым годом время, за которое объем создаваемых человечеством данных удваивается, становится все меньше, а рост производства носителей данных увеличивается не так быстро, и через несколько лет возникнет проблема хранения данных. Кроме того, обрабатывать все возрастающие объемы данных также непросто. Поэтому целесообразно накапливать и сохранять не «сырые данные», а некий результат их обработки, анализа и обобщения. Если данные – это не долго живущие новости, временные записи и т.п., не предназначенные для длительного использования, то информация представляет собой полуструктурированные (или агрегированные) данные, служащие, например, опорой для периодического принятия каких-либо решений.

В свою очередь знания, являющиеся результатом переработки информации, имеют весьма длительный цикл жизни, несут определенную идею и снабжены контекстом, определяющим область ее эффективного применения в данном месте в данное время. *Данные* – это объекты материального мира. *Информация* – знаки, полученные при преобразовании данных в сознании человека или в процессоре компьютера. Знания есть только у человека или в самом элементарном виде в информационной системе.

Управление знаниями (Knowledge Management) – это организация управленческих действий на базе всех информационных ресурсов фирмы [10]. При этом знания классифицируются и распределяются по категориям в соответствии с предопределенной, но развивающейся онтологией структурированных и слабоструктурированных баз данных и баз знаний. Цель системы управления знаниями – сделать знания доступными и повторно используемыми на уровне всей корпорации.

Иногда в BI включают технологию управления знаниями. Это связанные области информационных технологий, которые имеют немало отличий. Под управлением знаниями в общем случае понимается дисциплина, которая обеспечивает интегрированный подход к созданию, сбору, организации, доступу и использованию информационных ресурсов организации. Эти ресурсы включают в себя корпоративные базы данных, текстовую информацию, такую, как документы, описывающие правила и процедуры, и, что наиболее важно, неявные знания и опыт сотрудников организации.

Есть три основных компонента управления знаниями:

а) люди получают, генерируют и передают знания;

б) процессы используются для распространения знаний;

в) технологии обеспечивают быструю и эффективную работу людей и процессов.

Можно выделить следующие функции систем управления знаниями:

1. Сбор знаний.
2. Хранение и обработка знаний.
3. Доставка знаний.

Следует отметить, что KM больше ориентирована на анализ неструктурированной или слабоструктурированной информации (например, HTML), которая не является предметом анализа BI-инструментов. KM обеспечивает категоризацию, анализ и семантическую обработку текстов, расширенный поиск информации и др., а технология BI предназначена для анализа фактографической структурированной (базы данных, плоские файлы и другие ODBC или OLE DB-источники данных) и слабоструктурированной информации (например, XML).

Web-сервисы и Business Intelligence

Технология Web-сервисов [11] направлена на интеграцию приложений предприятия, призвана помочь BI достичь желаемых результатов, усилить его потенциал для правильного принятия решений и базируется на использовании точной информации в реальном масштабе времени, поступающей от всех уголков предприятия. Поэтому очень эффективным является использование технологии Web-сервисов в продуктах BI.

SOA позволяет связать различные системы, существующие на предприятии, и формализовать бизнес процессы их взаимодействия. В центре внимания SOA находятся не данные, а сервисы, которые являются бизнес-функциями, предназначенными для обеспечения согласованной работы больших, состоящих из множества частей, приложений. Технология SOA имеет очень хороший потенциал в отношении BI-систем. Она позволяет обеспечить прозрачный доступ к информации, собранной в «виртуальное» хранилище данных из различных операционных и аналитических источников в реальном времени.

Кроме того, **использование сервисов как основы для построения BI-системы** позволяет преодолеть многие трудности, связанные с клиент-серверной архитектурой [12]. Так, например, становится возможным управлять событиями, выполнять многие задачи в режиме реального времени, автоматизировать анализ и обработку информации, делать легко масштабируемые и «интегрируемые» системы.

Business Intelligence 2.0

Business Intelligence 2.0 – новое направление развития Business Intelligence. Это, подобно Web 2.0 и Semantic Web, скорее общая парадигма, абстрактная концепция, а не конкретное приложение [13]. Если в традиционном **Business Intelligence** анализируемая информация вовлекалась в решения, ориентированные на бизнес (оперативно обрабатывается поступающая информация и представляется в удобной пользователю форме) [14], то в **Business Intelligence** 2.0 информация помогает принимать решения еще до того, как произошло некоторое событие. **Business Intelligence** 2.0 поддерживается SOA-архитектурой, которая обеспечивает обратную связь принимаемых решений и информации, поступающей в режиме реального времени.

Основными особенностями технологий **Business Intelligence** 2.0 станут средства контроля производительности, интегрированного планирования и прогнозирования, встраиваемые в операции, аналитические порталы, рабочие пространства для коллективной работы и интеграция с MS Office, расширенные технологии поиска и визуализации данных, реализация функций BI в качестве сервисов, стандартизация и расширение применения открытых решений, развитие семантических методов, средства работы со слабоструктурированными данными и специализированные высокопроизводительные программно-аппаратные решения не только для хранилищ данных и задач BI.

Business Intelligence 2.0 обеспечивает более высокий уровень абстракции и работает с семантической моделью данных. Это позволяет осуществлять поиск нужной пользователю информации на семантическом уровне, используя ее метаописания. Для поддержки этого используются следующие технологии Semantic Web:

- Resource Description Framework (RDF);
- Web Ontology Language (OWL);
- SPARQL (SQL like query language for RDF).

Это обеспечивает новый уровень взаимодействия. Можно выделить шесть ключевых факторов, связанных с переходом BI на новый уровень.

1. Распространение BI за пределы отдельного предприятия через объединение оперативных и аналитических приложений.
2. Поддержка аналитической культуры и автоматизация принятия решений.
3. Готовность к взаимодействию с новым поколением работников, которое, в отличие от своих предшественников, вовлекает технологию BI в свою персональную жизнь.
4. Использование семантических технологий, использующих метаданные и онтологии для интеграции взаимодействия между людьми и процессами [15].
5. Адаптация стандартов и технологий Web 2.0 для совместной работы, масштабирования и ускорения создания BI-продуктов и приложений.
6. Достижение высокой производительности BI.

В табл. 1 представлены наиболее характерные явления в **Business Intelligence**, которые в ближайшее время будут реализованы в корпоративных программных комплексах.

Основные черты Business Intelligence 2.0. *Оперативность и сквозное взаимодействие; Встраивание в процессы; Качество обслуживания; Доступность.*

Таблица 1 – Основные возможности **Business Intelligence** 2.0

Всеобъемлющая BI-технология	Pervasive BI – расширение использования business intelligence во всем бизнес-сообществе за счет применения этой технологии среди сотрудников разных уровней.
Доступность BI для масс	Расширение диапазона BI-возможностей, доступных малому и среднему бизнесу и небольшим IT-отделам.
Ролевая BI-технология	Аналитические задачи, нацеленные на задачи и интересы конкретной аудитории. Показатели, метрики, тенденции, инструментальные и оценочные панели, предназначенные для конкретных функций (исследование, маркетинг, продажи, финансы и проч.) и на бизнес-уровне (стратегическом, тактическом, операционном).
Исследовательская аналитика	Интерактивные, исследовательские процессы обнаружения новых связей и явлений, когда каждый новый вопрос ставит дополнительные. Исследовательская аналитика проповедует принцип «прислушиваться к данным» и искать в них информацию.
Динамичная аналитика	Обеспечение быстрого отклика на различные ситуации, где необходим анализ. Подразумевает возможность быстро и эффективно адаптироваться к изменяющимся обстоятельствам как в бизнес-вопросах, так и в технических.

Business Intelligence 2.0 включает в себя несколько важных новых концепций, касающихся использования информации в бизнесе, организациях и правительственных структурах. Этот термин по сути своей связан с BI в реальном времени, технологией, управляемой внешними событиями, но основная идея состоит в применении этих методов к бизнес-процессам. **Business Intelligence** 2.0 – это расширенная технология, включающая динамическое и автоматизированное принятие решений.

Средства **Business Intelligence** нового поколения позволяют управляться событиями и смогут выявлять аномальные ситуации и возникающие проблемы, для этого в них будут встроены соответствующие интеллектуальные и адаптивные возможности. В качестве источников данных для анализа будут служить не только хранилища, но и данные самых разнообразных внутренних и внешних для компании источников и программных агентов. В технологиях нового поколения преобладающими

станут коллективные методы работы в сети с использованием виртуальных форумов, блогов, вики, социальных сетей. Реализация приложений будет опираться на SOA с применением общедоступных решений и активных элементов Web 2.0 (например, AJAX) и функционально богатого внешнего интерфейса.

Чтобы адаптировать бизнес-модели к современному миру, данные передаются в реальном времени через SOA, использующие слабосвязанные и обеспечивающие хорошее взаимодействие сервисы, которые стимулируют стандартизованную интеграцию приложений. В табл. 2 приведены свойства, которыми должен обладать Business Intelligence 2.0.

Таблица 2 – Свойства Business Intelligence 2.0

Управляется событиями	Автоматизированные процессы управляются событиями, поэтому очевидно, что для разработки более гибких процессов нужно анализировать и интерпретировать события. А значит, анализировать данные, от события к событию, как параллельно с бизнес-процессами, так и в форме отдельного этапа процесса.
Выполняется в реальном времени	Иначе нельзя реализовать BI-возможности в качестве этапа процесса, и тем более не удастся автоматизировать операции. Для сравнения: пакетные процессы – информационные, они содержат отчеты об эффективности процесса, но не могут быть частью самого процесса, за исключением тех случаев, когда время не играет решающей роли. Любое приложение, имеющее отношение к торговле, динамическому ценообразованию, оценке спроса, безопасности, оценки рисков, выявлению мошенничества, пополнению склада и любому взаимодействию с клиентом, – это процесс, зависящий от времени, а следовательно, требует обработки в реальном времени.
Автоматизированный анализ	Чтобы автоматизировать ежедневный процесс принятия решений, организациям необходимо не просто представить данные в виде инструментальной панели или отчета. Задача – превратить данные в реальном времени в нечто действенное. Необходимо их автоматически и динамически интерпретировать. На практике это подразумевает сравнение каждого отдельного события с тем, что ожидается в норме, исходя из прошлого опыта или прогнозируемой будущей эффективности. Продукты BI 2.0 должны в качестве основы использовать некоторую норму как на детальном, так и на агрегированном уровне, и сравнивать автоматически конкретные события с этой нормой.
Дальновидность	Чтобы понять, как данное событие повлияет на потребности организации, необходимо обладать некоторой дальновидностью. Чтобы ответить на вопросы: «Придет ли заказ вовремя?», «Откажет ли система сегодня?», – нужно уметь делать прогнозы. Эта возможность придает особое значение операционным отделам, которые должны представлять себе перспективу изменения эффективности своей работы в течение дня, недели или месяца.
Ориентация на процессы	Продукты BI 2.0 должны быть ориентированы на процессы. Это не значит, что процессы моделируются с помощью инструмента управления. Действия можно оптимизировать исходя из результатов конкретного процесса, хотя он может и не иметь точного определения.
Масштабируемость	Масштабируемость является краеугольным камнем BI 2.0. Потоки событий могут быть непредсказуемыми и очень интенсивными.

Выводы

В Украине рынок BI только начинает набирать силу и корпоративные пользователи все чаще задумываются о времени и средствах, которые можно сэкономить, используя эффективные бизнес-инструменты.

Интерес к системам BI приходит после того, как решены основные задачи, связанные с оперативным учетом на предприятии. К этому моменту в компании, как правило, работает ряд контуров одной или нескольких учетных систем, накапливающих информацию о совершенных операциях (транзакциях) в базах данных. С успехом, справляясь с оптимизацией оперативных задач, на которые эти системы ориентированы, они, однако, не позволяют увидеть общую картину деятельности компании.

Интерес к разработке продуктов BI в Украине постоянно возрастает. Например, украинский разработчик ПО Softline заявил о создании дочерней компании Softline BI, специализацией которой будет выполнение проектов внедрения деловых интеллектуальных систем BI.

Среди компаний, для которых сегодня актуально внедрение BI-систем, есть организации, традиционно тяготеющие к укрупнению, к инновациям. Например, финансовый сектор, где, с одной стороны, существуют жесткие требования государства к предоставлению отчетности, а с другой – растет конкуренция. Поэтому банки, страховые компании вынуждены применять системы, все более профессионально отвечающие требованиям, и технологии, которые позволяют принимать более обоснованные решения.

Например, IBM Cognos присутствует на территории стран СНГ с 1995 г. За это время ее заказчиками в регионе стали более 300 организаций. В Украине это: British American Tobacco Ukraine, ГК «Аптечный холдинг», холдинг Unitrade Group, «Международные Авиалинии Украины», медицинский центр «Добробут», ГК «Континиум Укрресурс», «Сумыоблэнерго», «Райффайзен Банк Аваль» (OTP Bank Plc.), Нефтяная компания «Альянс-Украина», «Ощадбанк», СК «Оранта», American Tobacco Ukraine, ГК «Континиум Укрресурс» и др. Все они являются клиентами компании Zept Group.

Проанализировав концепцию Business Intelligence и основные направления ее развития, можно сделать выводы о том, что это направление IT сейчас активно развивается, причем основные направления развития связаны с интеллектуализацией BI и их глобализацией, связанной с ориентацией на Интернет-технологии. Многие приложения BI становятся Web-ориентироваными, они представляются пользователям как Web-сервисы, а доступ к ним обеспечивается при помощи метаданных и онтологий.

Литература

1. Mehta, Sanjay. BI 2.0 Technology – MAIA Intelligence Perspective [Электронный ресурс]. – 2008. – Режим доступа : www.maia-intelligence.com.
2. Дэйв Вэллс. Десять основных преимуществ Microsoft Business Intelligence / Дэйв Вэллс; пер. с англ. – 2008. – Режим доступа : http://www.w3.org/TR/html4/strict.dtd.
3. Berners-Lee T. Weaving the Web.- Harper / Berners-Lee T. – San Francisco, 1999.
4. Enabling Technology for Knowledge Sharing / [Neches R., Fikes R., Finin T. и др.] // AI Magazine. – 1999. – Vol. 12, № 3.
5. Хью Уотсон. Современное состояние бизнес-аналитики (Business Intelligence, BI) [Электронный ресурс] / Хью Уотсон, Барбара Викском. – Режим доступа : http://www.osp.ru/
6. Ручкин К.А. Применение бизнес-интеллекта в страховых компаниях / К.А. Ручкин, С.Ю. Палиенко // Штучний інтелект. – 2005. – № 4. – С.137-139.

7. Скобелев П.О. Открытые мультиагентные системы для холонических предприятий / Скобелев П.О. // Штучний інтелект. – 2001. – № 3. – С.107-109.
8. Дэн Эверет (Dan Everett) Web-сервисы на службе у Business Intelligence [Электронный ресурс] / Дэн Эверет ; [пер. с англ. Intersoft Lab]. – 2007. – Режим доступа : http://www.iso.ru/journal/articles/572.html
9. Philipose Mathew. Web Services: A BI Perspective / Philipose Mathew, Rajarajeswari N., Visalachi Rameshkumar // Journal Patti Computer Systems. – Cambridge, (USA), October, 2006.
10. Semantic Web [Электронный ресурс]. – 2002. – Режим доступа : http://www.semanticweb.org.
11. Артемьев В. Что такое BUSINESS INTELLIGENCE? / Валерий Артемьев // Открытые системы. – 2003. –№ 4.
12. Маторин С.И. Применение теории паттернов в средствах «BUSINESS INTELLIGENCE» / С.И. Маторин, Д.Б. Ельчанинов, В.С. Маторин // Штучний інтелект. – 2004.– № 2. – С.142-153.
13. Гладун А.Я. Основи методології формування тезаурусів з використанням онтологічного та мереологічного аналізу / А.Я. Гладун, Ю.В. Рогушина // Штучний інтелект. – 2008. – № 5. – С.112-124.
14. Гладун А.Я. Применение тезауруса предметной области для повышения релевантности поиска в Интернете / А.Я. Гладун, Ю.В. Рогушина // Штучний інтелект. – 2005.– № 4. – С.742-753.
15. Гладун А.Я. Использование онтологических знаний и тезаурусов для объективного профилирования специалистов / А.Я. Гладун, Ю.В. Рогушина // Штучний інтелект. – 2006. – № 3. – С. 379-390.

А.Я. Гладун, Ю.В. Рогушина, Л.В. Петрухіна

Застосування технологій Semantic Web для керування знаннями в системах Business Intelligence

У статті розглянуто тенденції розвитку Business Intelligence та його зв’язок з іншими напрямками IT, аналізується доцільність застосування в Business Intelligence 2.0 технологій Semantic Web, зокрема онтологічного аналізу та стандартів представлення метаданих.

Anatoly Gladun, Julia Rogushina, Ludmila Petruhina

Application of Technologies Semantic Web for Knowledge Management in Systems Business Intelligence

The tendencies of development Business Intelligence, his communication with other directions IT are considered, the expediency of application in Business Intelligence 2.0 technologies Semantic Web, in particular of the ontological analysis and standards of representation metadata is analyzed.

Статья поступила в редакцию 01.04.2010.